

Rapid Word Collection, dictionary production, and community well-being

Brenda H. Boerger, SIL International
ICLDC5, March 2017

KEY WORDS: lexicography, fieldwork, community engagement

Rapid Word Collection (RWC) methodology collects words by semantic domain, usually during a two- to three-week workshop involving around 60 community members. I present evidence from Natqgu [ntu] to show that participation by community members in RWC workshops and the anticipated bilingual dictionary which results from it positively impacts their well-being, and positions RWC in particular and lexicography in general as high priorities for language development and community well-being.

1. Introduction

It has been observed that concomitant with language development in general there is an increase in the status and value of the language on one hand and the self-image of its speakers on the other (Ostler 2003:176, Boerger 2015:152). A dictionary is one product of language documentation and conservation which is frequently requested by communities. Some of the reasons dictionaries and their production increase well-being are among the following. First, participation in dictionary production increases language-related knowledge. Second, dictionaries are important for preserving language and culture data for posterity. Third, bilingual or trilingual dictionaries, especially, can be a conduit for understanding other cultures. Fourth, sometimes languages are not recognized as potential recipients of other governmental services until there is a dictionary or other language development product which confers status on the language. Fifth, a dictionary also serves as a literacy tool in the community and its schools, contributing to growing abilities in first language literacy.

VID: 01 Ben Metopz#3—use in school

Dictionary production also has advantages outside the community. Dictionary work benefits the scholar-fieldworker by generating an extensive wordlist which can help with the analysis of texts that have been collected, providing a foundation for text glossing and processing. Furthermore, RWC has the added benefit of facilitating the collection more words in less time than other methods, thereby increasing the effectiveness of fieldwork with regard to time invested, numbers of words collected, and native speaker consensus about the data (Boerger et al, in progress).

Assertions in this presentation are supported with data from a late 2015 RWC workshop in Natqgu [ntu]¹ spoken in the Solomon Islands, South Pacific. I led with a team of seven US interns as part of research for a Documenting Endangered Languages fellowship.² Following the workshop participants were invited to reflect about the workshop. No prompts or other guidance were given about what to share.

2. Stages and roles in a Rapid Word Collection workshop

Rapid Word Collection (RWC) workshops gather 20-60 members of a community together to collect words by sequencing through a list of semantic domains and subdomains. The

¹ Natügu [ntu] is spelled Natqgu in the local orthography. Since the local orthography was used for the RWC workshop, all Natqgu data will also be cited in that local orthography.

² The research reported on here was funded in part by the Documenting Endangered Languages (DEL) fellowship #FN-230212-15, entitled, “Natügu Dictionary and Legacy Texts.” DEL funding is cooperatively sponsored by US government agencies NSF and NEH. The author gratefully acknowledges this support.

benefit is that people are brainstorming in the vernacular and not normally a language of wider communication (LWC).

Producing dictionaries by means of RWC workshops can be viewed as having five stages: 1) administration and training, 2) word collection, 3) word annotation, 4) post-workshop wordlist cleanup, and 5) post-workshop editing of lexical database. More details can be found in Boerger et al, forthcoming.

Before the workshop takes place it is critical that two roles be filled: the **workshop coordinator** and the **logistics manager**. They are local community members who work together at the workshop site to gather people and resources for the workshop. The workshop **consultant** is usually an outside expert who leads the training sessions and assists the coordinator in giving feedback to the groups of participants. That is the role I filled in the Natqgu RWC workshop. A **record keeper** collects and files informed consent documents, records attendance, logs numbers of words collected by domain, and distributes a new domain folder to each group as they turn in completed folders. This role was filled by two of the seven US interns on my team.

Word collection is the primary focus of an RWC workshop and uses the most personnel. Six groups containing four to six fluent speakers of the language are required to finish the semantic domains in two weeks. Each group is composed of a **team leader**, a **scribe** who writes down the words, and from one to three **language experts**.

After the word collection groups submit their words, they get passed on to the word annotation personnel, the **glossers** and the **typists**. The glossers write the meaning of the vernacular word in the LWC and the typists add words to the Fieldworks Language Explorer (FLEX) lexical database using the “collect word” tool in the FLEX software to type both the vernacular word and its gloss. Following the workshop, one or two weeks are allotted to general cleanup of the words collected. This involves a few of the best qualified participants who edit the FLEX database to correct spellings, clarify glosses, and purge duplicates.

VID: 02 Ben Metopz#2—education, loss, glossing

Post-workshop clean up continues what was started following the workshop, with special attention to adding fully developed definitions, information about pronunciation, synonyms and antonyms, example sentences, part/whole and other lexical relations, dialect notes, parts of speech, grammatical appendices and any further information desired by the stakeholders, both the community and any linguists they relate to (Bartholomew and Schoenhals 1983).

3. Participation in dictionary production increases language-related awareness

Even though the participants were collecting written words for a dictionary, they were also learning more about language and linguistics at the same time. This is demonstrated by the two videos related to this section. One speaker acknowledges that language is primarily spoken and that they therefore need to speak it to their children in order to maintain it. The other one expresses the understanding that language change is natural and therefore it is normal for them to speak somewhat differently from people in the past.

VID: 03 Elizabeth Ilova—Ig is oral

VID: 04 Ben Metopz#1—Ig changes

4. Dictionaries preserve language and culture

In addition to the most commonly known lexical database functions of FLEX, there is also an anthropology notebook tab which allows the fieldworker to annotate by category according to the Outline of Cultural Materials (ref) and to add extensive notes. Since cultural information is also often endangered, it is important to document it while the expert practitioners

are still living. Items in the lexicon are tagged with semantic domain number categories through use of the “collect word” tool. Lexical items can also be linked to the anthropology notebook labels, thereby connecting the language and culture sides of the database. This provides a means of preserving and linking linguistic and anthropological information in the same database.

This was done by my intern who studied the process of weaving banana fibers using a backstrap loom. She added weaving craft terms to the lexicon using semantic domain categories, then provided further description and discussion of 17 weaving terms in the notebook tab.

Semantic Domains

4.3.9.1 - Custom
5.3.5 - Clothes for special people
6.6.1 - Working with cloth
6.6.1.1 - Cloth
6.6.1.2 - Spinning thread
6.6.1.4 - Weaving cloth
8.3.1.8 - Pattern, design

Anthropology Codes

280 leather, textiles, fabrics
285 mats and basketry
286 woven fabrics,
290 clothing
292 special garments.

In her entries about the banana fiber breechclout, the intern had the following entries. From the lexicon:

lrpz nqesa' NP 1 banana fiber breechclout; worn by men during nelc dance or by bridegroom at marriage ceremony 2 banana fiber dance dress worn by women for special occasions

Anthropology note in the lexicon tab:

Made from banana fiber. Formerly worn only by chieftains as it was more costly and labor-intensive to produce than the tapa worn by tribal members. Presently (2015) used by anyone who can acquire one (\$1000 SI) for the purpose of public ceremonies and dances. This garment is a visual reminder of a common cultural heritage shared by the four Santa Cruz languages: Natqgu [ntu], Nalrgo [nlz], Engdwu (formerly Nagu) [ngr], and Noipx [ISO code pending].

Anthropology note in the notebook tab:

I asked Selwyn Balq about who taught him to weave. Did he ask to be taught or was he forced to learn? His father told him he would learn to weave. When his father died, he realized the craft would die with him, so he set out to begin weaving. His first lrpz nqesa' was commissioned and he used one of his father's as a template to create the new one.

Participants in the workshop also recognized that there has been significant language loss due to contact with the national language, English, and the English-based creole, Solomon Islands Pijin, which is the language of wider communication in the country. The two speakers featured in this section talk about how the dictionary work is a way of preserving aspects of the language that would otherwise be lost, and acknowledged the role of the older speakers in the workshop as knowing more words than the younger speakers.

VID: 05 John-Steele Clq—preservation

VID: 06 Thomas Yade—loss preserve**5. Dictionaries are conduits for understanding other cultures**

While language documentation proponents agree that preserving language and culture is a worthy goal, they also acknowledge that languages do not live in monolingual vacuums. Often, then, one of the felt needs of communities involves access to products which help them function in a multilingual world. This is a normal byproduct of bilingual dictionaries, in that the definitions and usage notes serve to connect the cultures of the two languages. This was also recognized by the workshop participants, especially by one of the teachers who was able to come during a one-week break from school.

VID: 07 Levi Dobu—other cultures, lgs**6. Dictionaries confer recognized status on languages**

As more and more governments, NGOs, universities, and individuals are pursuing language documentation, competition for funds and for recognition are inevitable. The existence of a dictionary or even the granting of an ISO 639-3 code, therefore, can give substance to claims that variety X is actually a language, distinct from other varieties in a particular geographical area, and is therefore worthy of consideration for language development funding.

Hearing about a fourth related language on Santa Cruz Island, which speakers of the other three languages did not understand, provided the opportunity for interns to investigate more fully. Two of them collected a Swadesh 100-word list for that variety and as a result of submitting that work Noipä is now identified as a separate language with its own ISO 639-3 code [npx]. See also Boerger and Zimmerman (2012) which describes the process for Nalögo [nlz], another Santa Cruz language, which received an ISO code previously, and which provided impetus for a graduate student, whose fieldwork overlapped with ours, to study that language.

AUDIO FOR NOIPX OR NATQGU?**7. Dictionaries serve as community literacy tools**

A dictionary can also serve as a community literacy tool, contributing to growing competencies in literacy and taking its place in the schools and even on the smart phones of the community. This was confirmed at a March 2015 workshop for teachers sponsored by the Temotu Province Ministry of Education on Santa Cruz, Solomon Islands. The teachers were asked what the next step would be in moving toward multilingual education (MLE). Their response was that there needed to be a bilingual dictionary³ for Natqgu. Such a dictionary would supplement the existing literacy materials published from 1990 through 2005, which had been distributed to the primary schools in Natqgu-speaking areas. These literacy materials are enumerated in Table 1 (Boerger 2015:156).

Table 1. Natqgu Literacy Materials

| Date | Natqgu Title | English Title | Description |
|-----------------------------------|---------------------------|-----------------------------------|------------------------------------------------------------------|
| General Literacy Materials | | | |
| 1990 | Buk ngö Be Nöyö Më Natügu | Book of Stories Written in Natqgu | 17 stories produced in a writers' workshop by indigenous authors |

³ Personal communication, Elizabeth Ilovz, Kati Primary School teacher, Bznwz, Graciosa Bay, Santa Cruz, Solomon Islands.

| | | | |
|------|---------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|
| 1996 | Neke Aboole Bot? Dakxnzng Nan Trmclogo Be Ngr Lrkr' Ncdr Poi Inapi x Memwa' Nzvzng mz Skul Kuli kx Yrlqbrtile Poi Doa Ne Kio | Who Sank the Boat? The Crocodile's Food The Story of the Rat and the Pig Inapi and Memwa' Go to School The Dog that Tricked the Pig The Chicken's Child | 6 "Shell Book" easy readers, illustrated with line drawings ⁴ |
| 2002 | Be Kang Kqlu mz Natqgu | Many Stories in Natqgu | Graded reader with 43 titles: 17 from 1990, 6 stories from 1996, plus 20 more |
| 2004 | Buk kc Kai mz Nzrlwxngr mz Natqgu | The First Book for Reading in Natqgu | Alphabet primer for beginning readers; lessons authored by NLP team members |
| 2005 | (Review edition of Natqgu New Testament) | English-Natqgu, Natqgu-English wordlist | Wordlist requested by teachers and community |

During the same 15-year period, church-related materials and scripture translation materials were also published, providing a corpus of reading materials at all reading competency levels. Note, too that the 2005 bilingual wordlist provided some of the benefits of a dictionary until a more comprehensive one could be compiled.

These advantages were also recognized by participants in the RWC workshop, where one discussed learning to read and another learning to spell in the vernacular.

VID: 08 Henri Teti—learn to read

VID: 09 Susie Clq—relearning to write

8. RWC collects more words and senses in less time than corpus-based elicitation

There are numerous ways to approach collecting words for the production of a dictionary, including corpus-based and wordlist-based strategies (Boerger et al 2016). One method which shows great promise is the semantic domain-based Rapid Word Collection workshop methodology. This claim is supported in Boerger, Simons and Stutzman (In progress), where we look at 12 languages which conducted RWC workshops. Six of them were starting from scratch and six of them already had databases of a few thousand words. We found that on average, communities collected over 10,000 words, resulting in a minimum of 6,200 unique senses once words were edited following the workshop. Many languages had a higher percentage of unique senses. For Natqgu, we began with 5,500 senses going into the workshop. The community collected 13,856 words during the workshop, covering all of the semantic domains. This resulted in 11,500 total unique senses for Natqgu once the words were edited. This doubled the number of words collected during nearly 20 years of residence in the community, when dictionary production was not the focus. Rather, during that time, we only added new words to the lexicon as they came up in texts and in translation from English. Said another way, we doubled the output in 20 days from what we'd previously collected in 20 years.

Of the six languages which already had a lexical database in progress, Natqgu started with the highest number. The other languages, then, tripled or quadrupled the number of senses in their own lexicons. Meanwhile, the six languages without lexicons all collected well over 10,000 words, with anywhere from 3,700 to 12,500 resulting unique senses after editing.

By using the semantic domain strategy in the workshop setting we were able to collect multiple senses of a word, as well as multiple words to express a particular concept. These senses, synonyms and homophones are often missed during wordlist-based and corpus-based

⁴ The Shell Book RTU, PO Box 397, Ukarumpa, EHP, Papua New Guinea

elicitation. So, not only are more total words collected during RWC, but more depth of the lexicon and nuances of meaning are also more easily elicited using this strategy.

9. Dictionaries inform texts (and vice versa)

One of the texts we collected during 20 years in the Solomon Islands was the autobiography of a man who was born in 1921 and who died in 2012. He divided his hand-written manuscript into fifteen chapters with 81 section headings which came to 30 single-spaced pages once it was typed. This is the longest natively-authored written text in Natqgu and may be the longest written vernacular text in the Solomon Islands. It is significant for those reasons and for the history it provides alongside the day-to-day life of a man coming to maturity during WWII. The vignettes are full of drama, including the time he was hiding from soldiers in a hole during a rainstorm and the hole was gradually filling with water.

But getting back to our focus here, having a more complete wordlist following the RWC workshop has made the interlinearization of the text go more smoothly and quickly because most of the words needed are already in the dictionary. When there is a word in the autobiography which is not in the dictionary, it means one of several things: a) there is an omission and it needs to be added and glossed, b) there is a spelling mistake in the autobiography and once it is corrected it will gloss properly, or less often, c) this is a word from Nalrgo, Simon's first language, which he still uses when speaking Natqgu. So the words collected help with text glossing and the texts reveal possible gaps in the lexical database.

10. Conclusion

The evidence above shows that communities with dictionaries or lexicographic fieldwork benefit in five ways. It also demonstrates the usefulness of RWC as a fieldwork strategy. I conclude that the increased well-being recognized by the videos of community participants, in addition to the attested effectiveness of the methodology warrant incorporation of RWC workshops as part of best practice in lexicography for minority and endangered languages.

References

- Bartholomew, Doris A. and Louise C. Schoenhals. 1983. *Bilingual Dictionaries for Indigenous Languages*. Instituto Lingüístico de Verano, Mexico SIL. Accessed 20160509. <http://www.mexico.sil.org/resources/archives/2384>.
- Boerger, Brenda H., Gary F. Simons, and Verna Stutzman, In progress. Rapid Word Collection workshops: Community engagement yields more words in less time
- Boerger, Brenda H. 2015. Bible translation as Natqgu language and culture advocacy. In *Language Vitality through Bible Translation*. Beerle-Moor, Marianne and Vitaly Voinov, eds. Peter Lang: Berkeley Insights in Linguistics and Semiotics. 145-176.
- Boerger, Brenda H., Stephen Self, Sarah R. Moeller, and D. Will Reiman. 2016. *Language and Culture Documentation Manual*.
- Boerger, Brenda H. and Gabrielle Zimmerman. 2012. Recognizing Nalögo and Natügu as separate languages: Code-splitting in ISO 639-3. *Language and Linguistics in Melanesia* 30.1:96-133. <http://www.langlxmelanesia.com/>
- Ostler, Nicholas. 2003. Desperate straits for languages: How to survive. *Language Documentation and Description* 1:168-178.

Simons, Gary F. and Charles D. Fennig (eds.) 2017. *Ethnologue: Languages of the World, Twentieth edition*. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com> (accessed 20170226).

Wells, Kim Beebe. 2016. Rescued from Extinction: The Resurgence of Santa Cruz Banana Fiber Weaving. Dallas, TX: GIAL term paper.